

Public 32 Reporting

COVID-19 Case Surveillance Public Use Data Utility Summary

Users should consider the level of completeness, including suppression levels when planning their analyses and use of public datasets. Privacy protections will suppress field values to reduce reidentification risks. Completeness varies by jurisdiction (i.e., state, local, and territorial) and time period. Variables are consistently coded to the value “Unknown” when jurisdictions specify in the case data submitted to CDC that the value is unknown, the value “Missing” when jurisdictions do not provide a value, and the value “NA” when the value is suppressed as part of privacy protections.

Versions:

Dataset version: 5/3/2022

Quick Summary

	summary	all_fields_counts	all_fields_pct	quasi_fields_counts	quasi_fields_pct
1	total_rows	71,387,132	NaN%	71,387,132	NaN%
2	total_columns	32	NaN%	6	NaN%
3	total_cells	2,284,388,224	100.0%	428,322,792	100.0%
4	suppressed_fields	1,112,507	0.0%	1,108,587	0.3%
5	missing_fields	1,185,398,090	51.9%	52,732,720	12.3%
6	unknown_fields	192,123,731	8.4%	33,889,513	7.9%
7	non_blank_fields	905,753,896	39.6%	340,591,972	79.5%

Field Level Utility Summary

	variable	suppressed	suppressed_percent	missing	missing_percent
1	res_state	3606	0.0%	0	0.0%
2	res_county	3606	0.0%	0	0.0%
3	sex	115166	0.2%	186683	0.3%
4	age_group	46945	0.1%	749757	1.1%
5	race_ethnicity_combined	425493	0.6%	1720142	2.4%
6	hc_work_yn	513771	0.7%	50076138	70.1%
7	records_with_any_quasi_identifier	513771	0.7%	50155933	70.3%